AFRL-IF-RS-TR-2006-188
Final Technical Report
May 2006

EXPLORING THE UTILITY OF RESEARCH CYC FOR REASONING FROM NATURAL LANGUAGE

Leland Junior Stanford University

Sponsored by
Defense Advanced Research Projects Agency
DARPA Order No. S971/00

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

STINFO FINAL REPORT

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2006-188 has been reviewed and is approved for publication

APPROVED: /s/

PETER J. ROCCI, Jr.
Project Engineer

FOR THE DIRECTOR: /s/

JOSEPH CAMERA, Chief
Information & Intelligence Exploitation Division
Information Directorate

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
MAY 2006	Final	Sep 04 – Sep 05

4. TITLE AND SUBTITLE

EXPLORING THE UTILITY OF RESEARCH CYC FOR REASONING FROM NATURAL LANGUAGE

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA8750-04-2-0284

5c. PROGRAM ELEMENT NUMBER
62301E

6. AUTHOR(S)

Christopher Manning and Andrew Ng

5d. PROJECT NUMBER
COGV

5e. TASK NUMBER
00

5f. WORK UNIT NUMBER
05

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Leland Junior Stanford University
651 Sierra St.
Stanford CA 94305-5721

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/IFED
525 Brooks Rd
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-IF-RS-TR-2006-188

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; distribution unlimited. PA# 06-376

13. SUPPLEMENTARY NOTES
AFRL Project Engineer: Peter J. Rocci Jr., IFED, Peter.Rocci@rl.af.mil

14. ABSTRACT
This project investigated the potential for using ResearchCyc in natural language processing systems. The project focused particularly on natural language problems connected to sentence understanding, such as reading comprehension and robust textual inference. The project completed studies of the possibilities for using ResearchCyc knowledge for problems from this domain, it developed software for interaction between robust natural language processing systems and ResearchCyc, via its Java interface, and evaluated in a quantitative manner how much value could be gained from use of ResearchCyc and other alternative technologies by performing ablation studies.

15. SUBJECT TERMS
Knowledge bases, natural language processing, inference

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Peter J. Rocci, Jr.
U	U	U	UL	17	19b. TELEPONE NUMBER (*Include area code*)

TABLE OF CONTENTS

1.0 ABSTRACT

This project investigated the potential for using ResearchCyc in natural language processing systems. The project focused particularly on natural language problems connected to sentence understanding, such as reading comprehension and robust textual inference. The project completed studies of the possibilities for using ResearchCyc knowledge for problems from this domain, it developed software for interaction between robust NLP systems and ResearchCyc, via its Java interface, and evaluated in a quantitative manner how much value could be gain from use of ResearchCyc and other alternative technologies by performing ablation studies. The major results were: (i) despite its large size, in many places ResearchCyc still does not provide all the knowledge needed to reason about a problem, (ii) prospects are much better for using ResearchCyc to help with particular small pieces of a problem, (iii) for particular data sets that focused on knowledge-based problem solving, the results did show quantifiable gains from the use of ResearchCyc, but (iv) the gains from its use were often not large, but further research would be needed to assess to what extent this was due to limitations in the way the project employed ResearchCyc versus fundamental limitations in the content of ResearchCyc.

2.0 INTRODUCTION

This project was part of a larger effort interested in being able to provide robust, broad-coverage semantic understanding (Raina et al. 2005; Haghighi et al. 2005). In many situations it is clear that language understanding systems needs more "world knowledge" to be successful. The work of this grant explored the utility of Cyc (Lenat and Guha 1990), in particular, ResearchCyc, in improving the performance of such systems. We were interested in investigating the potential for using ResearchCyc in "bottom up" natural language processing systems. Specifically, the goal was to perform rigorous studies as to the extent to which the knowledge in ResearchCyc could add value to natural language processing (NLP) systems beyond the lexical and other knowledge already provided by other broad coverage resources such as WordNet.

The task of focus was an existing question answering and robust text inference system that uses a pseudo-logical representation of meaning for assessing answers and entailment. One key component of the system is that it has learned to incorporate knowledge from a number of diverse sources, such as knowledge regarding word semantic similarity, and word hypernymy/hyponymy. We built on this previous work and explored incorporating knowledge from Cyc into our text processing system. We explored how much leverage Cyc's knowledge base and reasoning can provide in this task, and compare it to the knowledge that can be obtained from less semantically rich sources such as WordNet or corpus-based semantic induction.

For example, given a sentence *The Israeli police arrested the robber*, we can automatically parse it in the logical representation Israeli [1] AND police [1] AND arrest [2 1 3] AND robber [3]. (I.e., that there exists an entity number 1, for whom the properties Israeli and police apply, that entity 3 is a robber, and that the arrest relation holds between entities 1 and 3.) Using the fact that arrest and catch often co-occur in documents and thus might be semantically related, our system is also able to make inferences such as that arrest[2 1 3] implies catch[2 1 3]. Using a logical theorem prover (e.g., Genesereth and Nilsson, 1987), this allows us to conclude that police [1] AND catch [2 1 3] AND robber [3]. I.e., that the police caught a robber. While the example described above was quite simple, we have successfully applied this system to solving robust textual inference problems. To construct a proof of each of the choices, the theorem prover typically has made certain assumptions (called abductive assumptions). For instance, in the example above, even though arrest and catch often co-occur, this does not conclusively imply that arrest has a semantic meaning closely related to that of catch. Thus, the inference that arrest [2 1 3] implies catch [2 1 3] is made at some cost, which reflects our uncertainty (more formally, the negative log probability) about the correctness of that inference step. Each proof typically requires multiple, small, assumptions, and the total cost of a proof is the sum of the costs of the individual steps. To answer a multiple choice reading comprehension problem, we then pick the answer that we were able to infer at the lowest cost.

One key component of our system is that it learns to incorporate knowledge from a number of diverse sources, such as knowledge regarding word semantic similarity, and word hypernymy/hyponymy. For example, WordNet actually tells us that arrest and catch are synonyms; thus, we might make the arrest/catch inference at significantly lower cost than if we knew only that they co-occur. More precisely, given a training set and external sources such as WordNet, word co-occurrence statistics, and so on, our system automatically learns how plausible assumptions such as arrest/catch are. More formally, it learns, using the external knowledge sources as features, what the costs of different inferences should be to give the most accurate possible QA system.

An error analysis of the mistakes made by our base system indicates that, on our corpus of multiple-choice reading comprehension questions, the vast majority of the errors (85%) were due to errors in generalized co-reference determination and/or due to lack of commonsense knowledge about the world. Lack of commonsense knowledge played a role in at least 70% of the errors made by our system. Since we already have methods for automatically combining multiple knowledge sources, we looked to incorporate Cyc as an extra information source for "commonsense knowledge". For example, our base system makes a mistake on a

problem where the passage talks the structure of the human heart/circulatory system, and the question asks about the regular beating of the heart. Because our system does not know that hearts beat regularly (unlike simpler inferences that can be made using WordNet), the correct answer was given a very high cost, and was not selected.

We conducted an extensive data analysis study to get a realistic picture of when ResearchCyc knowledge could and could not help us in such inference problems. Our conclusion was that while ResearchCyc sometimes has useful knowledge, there are many other situations in which knowledge remains incomplete and fragmentary.

3.0 USE OF RESEARCHCYC IN TEXTUAL INFERENCE SYSTEMS

For using ResearchCyc automatically in our existing and developing robust textual inference systems, we devoted considerable time to getting ResearchCyc working at all (it was very sensitive to exact Linux versions). We worked extensively with Cyc engineers in Austin to get ResearchCyc Natural Language tools up and running on Stanford's computers. We also devoted considerable time to understanding how to use ResearchCyc effectively (the learning curve is quite steep, and many areas of the natural language functionality and the Java interface to Cyc that we were attempting to use are not well documented). We built and tested a Cyc similarity module that utilizes taxonomic information in ResearchCyc. We ran our system end-to-end on the entire RTE development set and on data available from the ARDA AQUAINT program Knowledge-based Inference Pilot (KB Eval), both with and without our Cyc similarity module to produce quantitative experimental results detailing ResearchCyc's utility as a word-word directional similarity module in RTE. Results showed a reasonable positive impact in precision in our end-to-end system. Consider the change in proof cost between the older system and the system with ResearchCyc. We saw how the change correlates with the correct answer using the following statistic:

1 = always improves weights
0 = no value on average
-1 = always hurts us

After setting suitable weights for denotation, and known genls and isa relations for known concepts gives Correlation = 0.26

Thus Cyc has a reasonable positive impact in precision. The effect is not larger, and indeed Cyc helps on only a modest number of examples because of the sparseness of the usable information in Cyc.

Our term-similarity module was extended to uses several types of Cyc assertions (including multi-word strings, abbreviations, and denotations) to evaluate directional entailment. We also wrote and tested a module to get semantic translations of head verbs and roles given output from our dependency parser. We wrote a small system to test identify sub-categorization frames for verbs given a Stanford parse tree and suggest possible semantic translations of the head verb given these frames. It would be nice to have a system that could test different sub-categorization frames and then match their "event role representations to one another. We developed code to enable the assessment of event similarity across different surface semantic forms (in cases where sufficient sub-categorization frame mapping information was present in ResearchCyc). The system can test different sub-categorization frames for verb predicates and then match their "event role" representations to one another. This gives us a deeper form of event similarity: The system can assess whole verbal predicate/event similarity rather than working only with taxonomic similarities of individual nodes. Our Java system can currently get appropriate verb semantic translations in Cyc from a parse, and suggest events and roles played by verb argument according to this parse. But we did not have time to complete a phrase parser, and this in combination with sparse lexical coverage kept us from further progress in this direction. Based on

these extensions, we did further studies on the usefulness of ResearchCyc on the AQUAINT program KB Eval data. The outcome of these experiments was mixed. We achieved positive results for ResearchCyc helping with some data sets, particularly those that used simple English or focused more on "logical inference" type relationships (e.g., the PARC and Cyc data sets). There was no net positive value on some of the other data sets, which took complex sentences from real world contexts such as newswire (this includes Stanford's own data set). Overall, ResearchCyc improved our correlation between proof-cost and correct response; showing that these Cyc derived similarity scores were helpful and more informative than those of WordNet alone. The results of these studies are at present fairly contingent. They not only depend on how successful we have been at finding good ways to exploit ResearchCyc, but also on the nature of the problems and how sensitive our results are to different experimental conditions.

We also completed ablation studies on the performance of the system with and without various other components, as well as with ResearchCyc turned on. One result that we got that surprised us was how the system was fairly resilient to deleting components. This is perhaps reflective of the fact that the current system's performance level is quite modest, and it is always doing a lot of "guesstimate" reasoning, and it can do almost as well even with individual knowledge sources taken out.

4.0 USE OF RESEARCHCYC IN RECOGNIZING TEXTUAL ENTAILMENT

We have investigated the feasibility of using ResearchCyc as part of Stanford's system initially built for the PASCAL Recognizing Textual Entailment (RTE) challenge (see links below). The crucial question is the recall of ResearchCyc: how often is there sufficient taxonomic and reasoning information in ResearchCyc for it to be able to complete domain-independent natural language inference tasks.

We have chosen not to use ResearchCyc's parser or NL tools as we both were unable to successfully use CycNL components in the early releases of ResearchCyc, and were more interested in interfacing our NL tools' output with ResearchCyc Knowledge. Our plan is to parse sentences and identify grammatical relations using Stanford's tools, and garner information for inferring textual entailment using ResearchCyc's lexicon, argument-frame mapping, and concept hierarchy which can be plugged into various components of our system.

For readability, ResearchCyc's general "hash-dollar" relations (and those only) are italicized, so #$genls gets written as *genls*. All remarks and comments about what's there or not there are current as of ResearchCyc v1.0.

4.1 Relevant Links

The PASCAL Recognizing Textual Entailment Challenge
http://www.pascal-network.org/Challenges/RTE/

The Stanford system description: *Robust Textual Inference Using Diverse Knowledge Sources*
http://nlp.stanford.edu/~manning/papers/rte.pdf

Our parses and dependency analyses of the RTE dataset:
http://www.stanford.edu/~rajatr/rte/

ResearchCyc
http://researchcyc.cyc.com/

4.2 ResearchCyc Predicates

We rely on the ResearchCyc lexicon to interpret some or all of our dependency parses as conceptual information. There are several ResearchCyc predicates that make statements about relationships between concepts and words given sense, part of speech and/or sub categorization frame. The most basic of these is *denotation,* and the somewhat similar *semTrans* predicates generally contain more complicated conceptual representations of single words or groups of words. For example, *denotation* assertions look like:

(#$denotation #$Bat-TheWord #$SimpleNoun 0 #$Bat-Mammal)
(#$denotation #$Bat-TheWord #$SimpleNoun 1 #$BaseballBat)
(#$denotation #$Bat-TheWord #$Verb 0 #$BaseballBatting)

While a *nounSemTrans* assertion looks like:

(#$nounSemTrans #$Bachelor-TheWord 0
(#$and (#$isa :NOUN #$AdultMalePerson) (#$maritalStatus :NOUN #$Single))

and a *verbSemTrans* assertion (where word=*Feed-TheWord* sense=0, subcatFrame=*DitransitiveNPCompFrame):*

(#$verbSemTrans #$Feed-TheWord 0 #$DitransitiveNPCompFrame
(#$and (#$isa :ACTION #$FeedingEvent)
(#$fromPossessor :ACTION :SUBJECT)
(#$objectOfPossessionTransfer :ACTION :OBJECT)
(#$toPossessor :ACTION :INDIRECT-OBJECT)))

Denotations are generally useful in cases where proper translations of a word haven't been entered into ResearchCyc and something more cursory is acceptable (see ex. **6)**.
Crucially there are varying levels of granularity, relational, and definitional information in the lexicon, and each word can be "handled" by any number of such predicates.

The best documentation on ResearchCyc NL tools and lexical mapping is at http://www.cyc.com/cycdoc/ref/nl.html.

Once we are in the space of ResearchCyc concepts, a huge number of predicates ostensibly relate these concepts to one another. The *isa* (is a) and *genls* (generalization) predicates, which express hyper/hyponymy relations, are among the most reliably present and robust (the two most common, in fact, with 234,116 and 53,956 assertions, respectively). We have designed a search to explore the space of these relations to determine a path between two concepts in ResearchCyc as a measure of similarity/entailment in the common case where no direct connection exists. Also, the reader will see that the term "spec" gets used to talk about a specialization of a collection (the inverse of *genls*) , so *KidnappingSomeone* is a spec of *ActsCommonlyConsideredCriniminal.*

5.0 EXAMPLES FROM THE 2005 PASCAL RTE DATASET

All the following selections are "true" entailments from the Recognizing Textual Entailment dataset that our current system judged as "false", but which involve a relatively small piece of common sense knowledge not available in our baseline system, the knowledge of which is mainly what can be derived from WordNet. We've tried to choose entailment examples that are non-trivial, in that they involve some common sense or lexical "missing link" that our current strategy doesn't detect; and not *too* hard, in that they are straightforward and presumably require relatively few such links. They're listed by type and index, along with the non-trivial and missing link on our side of the inference, which is sometimes trimmed down for clarity.

We present ResearchCyc's capabilities and/or preventative deficiencies with respect to each entailment, and provide generalizations where possible. We also present brief discussions of the nature of general "Cyclish" relationships as they arise; this should be helpful and understandable for a potential user who hasn't mastered all terminology.

1. QA 591
T: *A jury is slated to decide for* the *first time whether Jack Kevorkian, famed as "Dr. Death," has violated Michigan's assisted-suicide ban, while the state continues to grapple with the issue of what to allow when the ill want to end their pain by ending their lives.*
H: *Jack* Kevorkian *is the real name of "Dr. Death".*
famed as Dr. Death -> is the real name of Dr. Death.
There is no entry for *Fame-TheWord* in ResearchCyc, so we are immediately prevented from working anywhere with this one. There is an entry for *Famous-TheWord,* which, as an adjective, denotes the concept *Famous*, but this knowledge wouldn't help us to understand the verb "fame".

2. QA 565
T: *Soprano's* Square*: Milan, Italy, home of the famed La Scala opera house, honored soprano Maria Callas on Wednesday when it renamed a new square after the diva.*
H: *La Scala opera house is located in Milan, Italy.*

Milan, home of the La Scala -> La Scala is located in Milan.

The Stanford Parser deals with the appositive well; the dependency output indicates that home is an appositive to Italy, that "La Scala Opera House" is an argument to the PP headed by "of", which in turn is an argument to the NP headed by home.
To recognize this entailment, we would want a module that translates the meaning of "home of" by identifying that the "home" noun phrase needs to be unpacked. We need to tell ResearchCyc that this instance of "home" corresponds to the *GenitiveFrame*, (i.e. part of a Genitive Phrase: nouns in association with a preceding possessive or a following 'of –PP') which in this case tells us that the *nounSemTrans* of *Home-TheWord* entails the *residesInDwelling* relation between two arguments to "home", or that POSSESOR "La Scala" resides in "Home" (the head NOUN). The appositive dependency informs us that "Home" in this case refers to "Milan".
In the hypothesis, we have that the *verbSemTrans* of *Locate-TheWord* entails the *objectFoundInLocation* between the verb's SUBJECT and OBJECT slots.
Given this coarse translation of both sentences, we look for relationships between the predicate-argument statements entailed by our translation. In this case, ResearchCyc does not have an obvious relationship between our crucially informative predicates: *objectFoundInLocation* and *residesInDwelling.*
We could conceivably leverage the *genTemplate* pred. in ResearchCyc, which generates more common and perhaps more statistically relevant English paraphrases given a template: in this case, telling us that "permanently located in" is the best paraphrase of *usualLocationOfObject*, the genlPred of *residesInDwelling*, and that "located in" is in fact the best paraphrase of *objectFoundInLocation.* We could imagine the general situation where words that aren't telling us much ("home" in this case) get semantic translations in ResearchCyc, from which we search the space of *genls (*generalizations) towards *semTrans' (*semantic translations) of target head verbs or predicates with matching arguments. In the case where no ResearchCyc relationship is clear (as above), we use *genTemplate* to translate back into English as we traverse and plug paraphrases back into Stanford's system at a cost as common paraphrase substitutions of less-informative predicates. Then again, there isn't an extensive paraphrase database and it's not clear how robust this method would be, but the sparsity would at least trim down the search space.

3. QA 594
T: *For* the *first time in history, the players are investing their own money to ensure the future of the game," Atlanta Brave pitcher Tom Glavine said.*
H: *Tom* Glavine *plays for the Atlanta Braves.*
Atlanta Braves **pitcher Tom ->**
Tom plays for **the Atlanta Braves.**
There are no baseball teams in ResearchCyc. Also, *Pitcher-TheWord* has only one denotation, and it's a *ServingVessel*, not a hurler.

4. IE 268

T: *There can be no doubt that the Administration already is weary of Aristide, a populist Roman Catholic priest who in December, 1990, won an overwhelming victory in Haiti's only democratic presidential election.*
H : Aristide *became president of Haiti in 1990.*
won victory **in presidential election -> became president**
Win-TheWord has only one *verbSemTrans* in ResearchCyc, as a transitive verb that takes one NP argument. There aren't translations for the *DitransitiveNP-PPFrame*, which is how we parse this example.
We easily find that "victory" is the object of the verb "win"; ResearchCyc would assert given our dependency tree that Aristide is a *winner-First* of some NP headed by "election". However, that's not clearly connected with the notion of "president" in ResearchCyc.

5. CD 674

T: *Jakarta lies on a low, flat alluvial plain with historically extensive swampy areas; the parts of the city farther inland are slightly higher.*
H: *The parts of Jakarta away from the coast are on slightly higher land.*

farther inland ->

away from the coast

The only ResearchCyc predicate containing some notion of "farther" is *fartherNorthThan*, and it is not related to any more generic lexical entries.

6. CD 801

T: *Reagan was* seriously *wounded by a bullet fired by John Hinckley Jr.*
H: *John W.* Hinckley *Jr. shot Reagan in the chest.*
The ResearchCyc entry for *Wound-TheWord* denotes an *IncurringAnInjury* action.
Unfortunately, the only *verbSemTrans* for *Shoot-TheWord* interprets the word as a denoting a *VisualImageRecording action.*
There is a *denotation* assertion stating that *ShootingAProjectileWeapon* is a concept denoted by the second sense of the verb *shoot.* However, there's no clear relationship, direct or indirect, between this predicate and *IncurringAnInjury.*

7. IR 102

T: *The White* House *failed to act on the domestic threat from al Qaeda prior to September 11, 2001.*
H: *White House ignored the threat of attack.*

Given the text sentence with correct dependencies, ResearchCyc would interpret "failed to act" in the *TransitiveInfinitiveVerbFrame* as denoting a *failureForAgents* relation between the subject "White House" and the action denoted by the INF-COMP "to act". (There is a multi-word string entry for ResearchCyc for "fail to make a payment", but none for "fail to act").
Also, *Ignore-TheWord* has no lexical information in ResearchCyc.

8. IR 36

T: *Scripps* Memorial *Hospital Encinitas emergency room doctors and nurses treat two to three injured surfers.*

H: *Scripps* Hospital *assists surfing accident victims.*
We've also got some good ideas about scoring verb similarity in a more informed (or at least different) manner then WordNet. Once we've identified the head verbs here (irrelevant of the NPs), we find a *verbSemTrans* for each and look at the *isa* for the head keyword: here *assist-TheWord* denotes a *HelpingAnAgent* action, with a *beneficiary* and a *performedBy* slot, and "treat "denotes a *MedicalTreatmentEvent* action, a direct hypernym, roughly, in ResearchCyc talk of *ServiceEvent*, again a direct hypernym of *HelpingAnAgent*. In addition to this *genl/genl* relationship, we see that both verbs have *performedBy* and *beneficiary* roles (two for two), indicating a not-entirely-superficial similarity that could be leveraged to score these verbs as close in meaning. One can see how these inferences could be easily generified as well.

9. IR 52

T: *Phish* disbands *after a final concert in Vermont on Aug. 15.*
H: *Rock band Phish holds final concert in Vermont.*
There's no lexical entry in ResearchCyc that has an appropriate translation for *Hold-TheWord,* namely the only *verbSemTrans'* in ResearchCyc corresponding to "holding" are *HeldCaptive*, *HoldingWithHand,* and *HoldingAnObject*, which specifies a physical object as an argument.
There is a *holds*-Underspecified relation that specifies a generic "holding" relationship, but this appears too complicated to be treated as a case in general.

10. IR 64

T: ***The wait time for a green card has risen from 21 months to 33 months in those same regions.***
H*: It takes longer* to *get a green card*
ResearchCyc has a good semTrans for *Rise-TheWord* denoting an *IncreaseEvent* action, where the *objectActedOn* is in the subject position. The construction "It takes longer" here is a tricky one, though; I don't see any way that ResearchCyc could interpret this usefully.

11. IR 79

T: *The privately owned spacecraft only got about 400 feet into space, according to radar measurements, but it was* enough *to confirm that it no longer takes a well-heeled government project to organize space travel.*
H: *private spaceship launches.*

The single verbSemTrans of *Get-TheWord* requires an ADJP complement, where it denotes an IntrinsicStateChangeEvent: the object of state change (the SUBJ) is the argument of (*toState SUBJ ADJ).* *Into-TheWord* has only *prepSemTrans*' in the *VerbPhraseModifyingFrame.* It's not clear that we could reconcile these expected differences (ADJP complement vs. VP modifier).
It's also the case that the only *verbSemTrans* for *Launch-TheWord* requires an NP complement.

12. MT 1228

T: *An official of* Abyan *police, where 16 Western tourists are being held since yesterday, announced that the hostages are held by the Yemeni "Islamic Jihad" group, which is demanding the release of its leader and lifting the embargo on Iraq.*
H: *The Yemen branch of the "Islamic Jihad" group, kidnapped the 16 Western tourists.*

hostages are held by the Yemeni group -> the Yemeni group kidnapped the tourists

Kidnap-TheWord is translated as a *KidnappingSomeone* action with a *perpetrator* slot.
KidnappingSomeone is a spec of *ActsCommonlyConsideredCriminal*, *TakingAPersonPrisoner,* and *CriminalAct.*
There is a compoundString entry for "hold hostages" that denotes a *HoldingHostages* action. *HoldingHostages* is a spec of *ActsCommonlyConsideredCriminal* and *HostileSocialAction.* Aside from knowing that both actions are criminal, ResearchCyc doesn't connect this with more specificity.

13. PP 487

T: *Located* just *three miles from Tullamore and only 45 minutes from the K Club, venue of the 2006 Ryder Cup, is Esker Hills, a genuine hidden gem and one of Irish golf's best kept secrets.*
H: *The K Club will host the 2006 Ryder Cup.*

K Club, venue of the Ryder Cup -> K Club will host the Ryder Cup
ResearchCyc provides us here a translation of "venue" to *eventOccursAt* with the correct filler slots. The translation of *host-TheWord* in the verb frame invokes a *hostOfEvent* action, though there's not a clear connection between this and *eventOccursAt.*

14. QA 1454

T: *In fact, Woolsey had had no first-hand experience with the world of spies until President Bill Clinton appointed* him *Director of Central Intelligence.*
H: James Woolsey is the director of the CIA.
The only relevant translation of *appoint-TheWord* invokes an *AppointingAmbassador* action. The assertions at this level are specific to "The collection of events where a state appoints an ambassador to another state" and can't handle the sort of event analysis of a "director appointment" required to help on this entailment.

15. CD 693

T: *This* growth proved *short-lived, for a Swedish invasion (1655-56) devastated the flourishing city of Warsaw.*
H: *Warsaw* was *invaded by the Swedes in 1655, and the city was devastated.*
Swedish invasion **devastated Warsaw -> Warsaw was invaded by Swedes**
Identifying that the "city was devastated" is straightforward here: we have a direct match in the dependency parse. The tricky issue in this example is how to understand the first half of the hypothesis. ResearchCyc identifies that "invasion" is the singular form of *Invade-TheWord,* which denotes a *MilitaryInvasion.* Unfortunately, there isn't any lexical information about how to translate *Invade* in ResearchCyc, for example a statement about a how an invasion may have a "performedBy" slot; *Invade-TheWord* currently has no lexical assertions at all. This hinders us from using ResearchCyc to our advantage in this entailment.

16. CD 735

T: *Even* more *than other economic activities, Mexico's financial services are concentrated in the capital.*
H: *Industry, retail stores, finance, and communications are all centered in the capital.*

Financial Services -> Industry, retail stores, finance, and communications
This example would require our system to do a particular sort of noun phrase matching order to identify each enumerated noun was a type of "Financial Services". This sort of noun-phrase entailment is something that would be generally useful were it robust; our system already implements a 'NP-match' function which currently relies only on WordNet. Such a function could conceivably query ResearchCyc to determine better-defined hypernymy in this case identifying that the hypothesis subject NP is composed of parts all of which are conceptual instances of a matching slot present in the text.

The only translation for the text NP here is the ResearchCyc assertion that the multi-word string "Financial Services" denotes a *FinancialCompany,* which in this case is not a hypernym of "finance" (which denotes FinancialOrganization*)* or "communications" (which doesn't have an appropriate lexical entry, the only one has to do with *CommunicationEffectiveness).*

17. CD 767

T: Hepburn*, a four-time Academy Award winner, died last June in Connecticut at age 96.*
H: Hepburn*, who won four Oscars, died last June aged 96.*
ResearchCyc has got no entry for "Academy Award" or "Oscar".

18. CD 779

T: *Voting for a new* European *Parliament has been clouded by apathy.*
H: *Apathy clouds EU voting.*

European Parliament : EU
ResearchCyc relates the concept *EuropeanUnion* to the string "EU" with the *initialismString* predicate, (a special case of acronymString where the string is formed using the first letters of the constituent words. *InitiliasmString* connects over 500 abbreviations to concepts in ResearchCyc). We could replace this string in the parse, and our similarity measure would here assign a higher score matching "European Union" to "European Parliament" than if we had only had "EU".

19. CD 820

T: *Kessler's team* conducted *60,643 face-to-face interviews with adults in 14 countries.*
H: *Kessler's team interviewed more than 60,000 adults in 14 countries.*

Conducted interviews = interviewed
We've got machinery in place to handle the "more than 60,000" = 60,643 equivalency. What is crucially missing on our side is the bolded equivalence above. The only appropriate lexicalization of the *Conduct-TheWord* would insist that "Kepler's Team" was the *directingAgent* of some ACTION called "interview". *Interview-TheWord* has an *agentiveNounSemTrans* that looks like this:

```
agentiveNounSemTrans Interview-TheWord 0 GenitiveFrame
    (and
      (interviewee ?ACT :POSSESSOR)
      (interviewer ?ACT :NOUN)))
```

It's not clear to me whether this is appropriate for a translation of the instance above.

20. IR 128

T: *Hippos do come* into *conflict with people quite often.*
H: *Hippopotamus attacks human.*

ResearchCyc actually has an entry for *Hippo-TheWord* linking the occurrence to the concept *Hippopotamus.* What remains then is to understand the "come into conflict" construction. The lexical information for *conflict-TheWord* is sparse, and there aren't any *multiWordString* definitions that capture the meaning of this construction.

A Toy Example

We construct this simple but non-trivial example as a starting point to help us understand what machinery needs to be in place to get arguments aligned and translated to meaningfully related ResearchCyc concepts. It's worth walking through how we'd set up a ResearchCyc query to ask about lexicalizations given typed-dependency output.

T: *John bought a car from Paul.*

F: *Paul sold a car to John.*

The Stanford parser gives us these typed dependencies from the sentences:
T: nsubj(bought, John) **det**(car, a) **dobj(**bought, car) **from**(bought, Paul)

F: nsubj(sold, Paul) **det(**car, a) **dobj(**sold, car) **to(**sold, John)

Our system identifies the head verb of each sentence and attempts to find the best fitting *subcatFrame* for each given the dependency parse. In this case, we see that the head verb in both cases has two complements, and NP and a PP.

Starting with the Text sentence, we check to see if there is a *DitransitivePP-NP* frame for *Buy-TheWord*. There is not. Then we look to see if there is an equivalent translation, in this case using ***(PPCompFrameFn*** DitransitivePPFrameType ***From-TheWord)*** to represent the subcat frame where there is an NP and a PP argument headed by "from". It's my understanding that this redundancy exists because the *DitransitivePP-NP* captures translations that are valid independent of the preposition, whereas the latter places constraints on the preposition and correspondingly the semantic translation. The ResearchCyc translation tells us that the following are true:

(isa :ACTION Buying)

(seller :ACTION :SUBJECT)

(objectPaidFor :ACTION :OBJECT)

(buyingPerformer :ACTION :OBLIQUE-OBJECT)))

On to the hypothesis,: we check to see if there is a *DitransitivePP-NP* frame, there is not. We check to the PPCompFrame as above and find the following translation:

(verbSemTrans Sell-TheWord 3

(PPCompFrameFn DitransitivePPFrameType To-TheWord**)**

(and

(isa :ACTION OfferingForSale)

(performedBy :ACTION :SUBJECT)

(transferredObject :ACTION :OBJECT)

(target :ACTION :OBLIQUE-OBJECT)))

In fact this is a problem. We would have got the correct translation (which corresponds to a *Buying* action, not an OfferingForSale action) only if we had asked for the *TransitiveNPFrame*. Unless we explicitly tried both, we've introduced a mistranslation which isn't necessarily recoverable, because the relationship between *OfferingForSale* and *Buying* isn't well-defined in ResearchCyc.

6.0 CONCLUSIONS

We present some summary remarks on the utility of ResearchCyc in recognizing textual entailment using our NL.

Lexical Coverage

To be sure, lexical coverage is the deficiency in ResearchCyc which hurts us the most on this task, and it is especially problematic in the absence of functional ResearchCyc NL tools. In most cases we find sparse or suboptimal lexicalizations that render any further search useless. Even on our toy example, the absence of a proper translation for "sells *X* to *Y*" keeps us from making the meaningful connection that we would expect from ResearchCyc: that both verbs express a buying action and can be translated as such given their NP-PP arguments.

True, we can implement searches that traverse the space of ResearchCyc relations and probably get some utility even if we have mistranslated the verb, but we would hope that for most examples that the right translation is in the KB: even too many ambiguous translations would be better than none.

Concept Linkage

It is hard to discuss this in general given the expert nature of much of ResearchCyc's knowledge, but for our purposes the concept linkage is also lacking in most examples: empirically speaking, we can almost never get from one sentence to the other using ResearchCyc alone. To this extent, ResearchCyc as a standalone RTE system is currently infeasible. Word-level similarity modules (that tell is that "hippo" means "hippopotamus", or that a "mosque" is a "building", or that "EU" designates the European Union), however, may be generally useful even in the situation where ResearchCyc can't handle arbitrary lexical lookups and conceptual connections. We intend to further explore using ResearchCyc for such similarity calculations in future work.

7.0 REFERENCES

Aria Haghighi, Andrew Y. Ng, and Christopher D. Manning. 2005. Robust Textual Inference via Graph Matching. HLT-EMNLP 2005, pp. 387-394. http://nlp.stanford.edu/~manning/papers/rte-emnlp05.pdf

Rajat Raina, Andrew Y. Ng, and Christopher Manning. 2005. Robust textual inference via learning and abductive reasoning. AAAI 2005, pp. 1099-1105. http://nlp.stanford.edu/~manning/papers/aaai05-learnabduction.pdf